# Sequence Specific Collective Motions in a Winged Helix DNA Binding Domain Detected by $^{15}$N Relaxation NMR†

Changwen Jin, Ian Marsden,‡ Xiaoqun Chen, and Xiubei Liao*

*Department of Biochemistry and Molecular Biology, College of Medicine, M/C536, University of Illinois at Chicago, 1853 West Polk Street, Chicago, Illinois 60612-4316*

*Received January 6, 1998; Revised Manuscript Received February 25, 1998*

ABSTRACT: The recognition between transcription factors and their DNA binding sites is a highly dynamic process. During transcriptional regulation, transcription factors must bind to or dissociate from their cognate DNA binding sites. The winged helix DNA binding motif is one of many highly conserved DNA binding motifs identified in transcription factors. Backbone dynamics has been studied on the $^{15}$N- and $^{2}$H-enriched winged helix family member Genesis. Our data show that the overall motions of the single domain Genesis are better described by more than two autocorrelation times ($\tau_m$). Our data also demonstrate that Genesis shows structure specific conformation exchange characterized by $R_{ex}$. Therefore, our results indicate that the structure of Genesis is highly dynamic and that secondary structure elements in Genesis have collective motions in the nanosecond to millisecond time scale. Since the winged helix DNA binding motif is highly conserved, this unique dynamic property observed in Genesis is also likely to be conserved in other winged helix family members and important in DNA binding.

Transcription factors play critical roles in cell growth, development, and differentiation by binding to DNA control elements and regulating gene expression. To understand how transcription factors control gene expression, the structures of many transcription factors and their DNA complexes have been studied. In these studies, several highly conserved structural motifs have been identified, and the atomic basis for sequence specific DNA recognition used by these motifs has been determined by NMR[1] and X-ray crystallography (*1*−*3*). However, these static structural methods give less information on how the DNA binding process proceeds as a protein and its cognate DNA site bind to form a complex. This process is expected to be highly dynamic. Proteins in a structurally conserved DNA binding motif may also follow a conserved dynamic pathway to recognize their DNA binding sites.

One of the highly conserved DNA binding motifs is the winged helix motif, which is found in a large family of DNA binding proteins that play important roles in cell function (*4*). This motif was initially identified in the mammalian hepatocyte-enriched transcription factor HNF-3 (*5*) and the *Drosophila* fork head homeotic protein (fkh) (*6*). Later, many winged helix proteins were identified in organisms ranging from yeast to humans (*7*−*16*). Previous studies have demonstrated that HNF-3/fkh homologues play important roles in developmental and tissue specific gene regulation (*17*−*22*). Previous structural studies also indicate that this DNA binding motif contains three or four helices, three sheets, and two wings (*23*, *24*), of which helix 3 makes most of the contacts in the major groove of DNA. Furthermore, several experiments showed that DNA sites are bent in winged helix protein−DNA complexes (*4*, *25*). These data indicate that a winged helix protein recognizes a linear DNA site first and then changes its conformation to form a stable protein−DNA complex. However, how the winged helix DNA binding motif can support this structure change is not well understood. It seems the dynamic property of the winged helix motif should play a role in this process.

To address this question, we studied the internal motion of one winged helix protein, the product of the Genesis gene, formerly HFH-2 (*11*, *26*), by using $^{15}$N relaxation NMR techniques. Genesis has been shown to be specifically expressed in embryonic stem cells or their malignant equivalent, and in addition, cotransfection experiments revealed that Genesis is a transcriptional repressor (*26*). In this report, we demonstrate that Genesis is highly dynamic and has collective motions among helices, sheets, and wings on a time scale of nanoseconds to milliseconds. This collective motion is likely to be a unique dynamic property associated with the

† This work was funded by the National Institutes of Health (Grant GM 52034 to X.L.) and the American Heart Association (Grant 95010100 to X.L.). X.L. is recipient of JFA awards from the American Cancer Society. The Bruker DRX600 instrument was purchased with funds from the University of Illinois at Chicago and a grant from the NSF Academic Research Infrastructure Program (BIR-9601705). Preliminary studies also made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH Grant RR02301 from the Biomedical Research Technology Program, National Center for Research Resources. Equipment in the facility was purchased with funds from the University of Wisconsin, the NFS Biological Instrumentation Program (DMB-8415048), the NFS Academic Research Program (BIR-9214394), the NIH Biomedical Research Technology Program (RR02301), the NIH Shared Instrumentation Program (RR02781), and the U.S. Department of Agriculture.

* To whom correspondence should be addressed.

‡ Current address: Abbott Laboratories, 1401 N. Sheridan Rd., Chicago, IL 60064.

[1] Abbreviations: HNF, hepatocyte nuclear factor; fkh, forkhead; NMR, nuclear magnetic resonance; $T_1$, longitudinal relaxation time constant; $T_2$, transverse relaxation time constant; NOE, nuclear Overhauser effect.

S0006-2960(98)00031-2 CCC: $15.00 
Published on Web 04/08/1998

winged helix DNA binding motif and may play an important role in protein−DNA interaction.

## MATERIALS AND METHODS

*Expression and Purification of the DNA Binding Domain of Genesis.* The gene encoding the DNA binding domain of Genesis (*27*), which in this report will henceforth be referred to only as Genesis, was generated by PCR amplification from the rat genomic clone (*11*). The endonuclease recognition sites *Nde*I and *Xho*I were engineered for the cloning of Genesis into pET21b (Novagen) with a 6XHis tag fusion at the C terminus. Therefore, the expressed protein contains the functional DNA binding domain of Genesis with an extra Met at the N terminus and Leu-Gln-$(His)_6$ at the C terminus for purification purposes. One of the problems in NMR studies of Genesis is that the resonances are highly overlapped. Thus, to reduce line width in the $H^N$ dimension and increase signal intensities, perdeuteration was carried out to replace carbon-attached protons. The protein was produced in *Escherichia coli* strain HMS174 grown in 98% $D_2O$ by induction of T7 polymerase at midlogarithmic growth through the addition of IPTG to 1 mM. Uniformly labeled Genesis was grown in isotopically enriched minimal media containing 0.6% $Na_2HPO_4$, 0.3% $KH_2PO_4$, 0.15% NaCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.5% Basal Medium Eagle Vitamin Solution (Gibco), and a rare element supplement. $(^{15}NH_4)_2SO_4$ (1 g/L) and 4 g/L unlabeled glucose were used for $^{15}N$ labeling. Protein expression was induced at an $OD_{600}$ of 0.4 in $D_2O$, and cells were grown for a further 24 h with the container open to air. Since the majority of expressed Genesis was present in insoluble inclusion bodies, a standard procedure using denaturing conditions was performed to extract the protein in $H_2O$ with Ni-NTA resin [*Qiagen Manual* (1992) 2nd ed.]. After elution of Genesis in 8 M urea at pH 4.5, the protein was renatured by dialysis against 100 mM phosphate buffer at pH 6.5 in $H_2O$. Thus, renatured Genesis is fully protonated on all amides.

Purified Genesis was exchanged into NMR buffer [50 mM phosphate (pH 6.5), 100 mM NaCl, and 10 mM $Na_2S_2O_4$ in 10% $D_2O$/90% $H_2O$] by ultrafiltration. Due to precipitation during dialysis and ultrafiltration, the final concentration for NMR studies was about 1 mM.

*Gel Shift Assay.* To study the isotopic effects in the Genesis−DNA interaction, gel shift assays (*28*) were performed to estimate the dissociation rate of $^{15}N$-enriched Genesis−DNA complexes and $^{15}N$-, $^{2}H$-, and $^{13}C$-enriched Genesis−DNA complexes. $^{15}N$-enriched Genesis or $^{2}H$-, $^{13}C$-, and $^{15}N$-enriched Genesis was incubated with $^{32}P$-labeled Genesis specific DNA probe #HFH-2#12 (*27*) for 30 min at 30 °C prior to the addition of a 25-fold molar excess of unlabeled DNA probe. Binding reactions were performed in 20 $\mu$L of 1× gel shift buffer [20 mM HEPES (pH 7.9), 40 mM KCl, 2 mM $MgCl_2$, 1 mM DTT, and 4% Ficoll] containing 50 ng of protein, 1 ng of labeled DNA, 2 $\mu$g of poly(dI-dC), and 200 ng of bovine serum albumin before adding 25 ng of unlabeled DNA probe. After incubation for 0, 10, 20, 40, and 60 min with unlabeled probe at 30 °C, samples were analyzed on a 9% nondenaturing polyacrylamide gel.

*NMR Method.* $^{15}N$ $T_1$, $T_2$, and $\{^1H\}-^{15}N$ NOE spectra were acquired on a Bruker DRX 600 MHz spectrometer equipped with a triple-resonance three-axis pulse field gradient probe. The pulse sequences used to record $^{15}N$ $T_1$ and $T_2$ and the steady-state $^1H-^{15}N$ NOE were published previously (*29*) and, therefore, will not be repeated here. For coherence selection gradients, the gradients on both the *x* and *z* directions were applied with the *x* gradient:*z* gradient ratio equal to 1.8 to achieve the best water suppression. $^{15}N$ $T_1$ relaxation parameters were measured from the spectra recorded with different durations of delay *T*: $T = 0.01$, 0.01, 0.1, 0.3, 0.5, 0.6, 0.7, and 0.9 s. $T_1$ spectra were recorded with magnetization relaxation as $\exp(-t/T_1)$ and in such a way that the delay between scans affected only the sensitivity and not the extracted $T_1$ values (*29*, *30*). $T_2$ values were determined from spectra recorded with delays *T* of 7.8, 7.8, 24, 39, 55, 71, and 95 ms. $^1H$ 180° pulses were inserted during the *T* relaxation times to eliminate the effects of cross relaxation times as described previously (*31*−*33*). $^1H-^{15}N$ steady-state NOE values were determined from spectra acquired in the presence and absence of a proton presaturation period of 3 s. $^1H$ saturation was achieved with the use of a 120° $^1H$ pulse at 5 ms intervals (*34*). Relaxation delays of 1.1 s were employed in measurements of $^{15}N$ $T_1$ and $T_2$ relaxation parameters. Relaxation delays of 2 s were used for the $\{^1H\}-^{15}N$ steady-state NOE experiment with $^1H$ saturation, and 5 s was used for the experiment without $^1H$ saturation.

*Determination of Relaxation Parameters*. The intensities of the peaks in the two-dimensional (2D) spectra were determined by the peak-picking package in the commercial software SYBYL (Tripos Inc.). The relaxation rate constants were determined by fitting the measured peak heights to a two-parameter function

$$I(t) = I_0 \exp(-tR_i) \tag{1}$$

where $I(t)$ is the intensity after a delay of time *t* and $I_0$ is the intensity at time 0. $R_i$ represents the spin−lattice ($i = 1$) and spin−spin ($i = 2$) relaxation rate constant, respectively.

The peak intensities were fitted to a single-exponential decay function by the nonlinear least-squares method as described previously (*35*). The steady-state $\{^1H\}-^{15}N$ NOE enhancements were calculated as ratios of the peak heights in the spectrum recorded with proton saturation to those in the spectrum recorded without saturation.

$$\mathrm{NOE} = I_{\mathrm{sat}}/I_{\mathrm{unsat}} \tag{2}$$

The standard deviation of the NOE value $\sigma_{\mathrm{NOE}}$ was estimated on the basis of measured background noise levels as

$$\sigma_{\mathrm{NOE}}/\mathrm{NOE} = (\sigma_{\mathrm{sat}}/I_{\mathrm{sat}}) + (\sigma_{\mathrm{unsat}}/I_{\mathrm{unsat}}) \tag{3}$$

where $I_{\mathrm{sat}}$ and $I_{\mathrm{unsat}}$ are the intensities of a resonance in the presence and absence of proton saturation, respectively. $\sigma_{\mathrm{sat}}$ and $\sigma_{\mathrm{unsat}}$ represent the standard deviations of corresponding intensities determined from the root-mean-square noise of background regions, respectively. As suggested previously, the standard deviation for each peak can be estimated by this method (*29*).

*Determination of Dynamic Parameters.* Amide $^{15}N$ nuclear spin relaxation is dominated mainly by dipolar interaction with its directly attached protons and by chemical shift anisotropy (*36*). The $^{15}N$ $T_1^{-1}$ and $T_2^{-1}$ relaxation rates and

the $^1H-^{15}N$ steady-state NOE can be expressed as a function of spectral density $J(\omega)$ evaluated at up to five different frequencies and an added term $R_{ex}$ in the expression for $T_2^{-1}$ incorporated to describe the effects that mainly result from the conformation exchange averaging (*37*).

$$1/T_1 = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N) \quad (4)$$

$$1/T_2 = 0.5d^2[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + {}^1/_6c^2[3J(\omega_N) + 4J(0)] + R_{ex} \quad (5)$$

$$\mathrm{NOE} = 1 + (\gamma_H/\gamma_N)d^2[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]T_1 \quad (6)$$

where the constants $c^2$ and $d^2$ are defined as

$$c^2 = (2/15)\gamma_N{}^2H_0{}^3(\sigma_{||} - \sigma_\perp)^2 \quad (7)$$

$$d^2 = 0.1\gamma_H{}^2\gamma_N{}^2h^2/(4\pi^2)(1/r_{NH}^3)^2 \quad (8)$$

where $\gamma_H$ and $\gamma_N$ are the gyromagnetic ratios of the $^1H$ and $^{15}N$ nuclei, respectively, $\omega_H$ and $\omega_N$ are the Lamor frequencies of $^1H$ and $^{15}N$, respectively, $r_{NH}$ is the internuclear $^1H-^{15}N$ distance for amides ($\sim$1.02 Å), $H_0$ is the magnetic field strength, and $\sigma_{||}$ and $\sigma_\perp$ are the parallel and perpendicular components of the $^{15}N$ chemical shift tensor, respectively. As described previously (*38*), $\sigma_{||} - \sigma_\perp$ is $-160$ ppm for the amide NH.

The NMR relaxation parameters $T_1$ and $T_2$ and the $^1H-^{15}N$ steady-state NOE are functions of the internal motions and autocorrelation time ($\tau_m$) of the molecule at a given field strength. These relaxation parameters are related to molecular motions through the spectral density function, and in the original and extended forms of the model free approach (*39*−*42*), the spectral density functions of an isotropically tumbling molecule are expressed as follows.

$$J(\omega) = S^2\tau_m/(1 + \omega^2\tau_m{}^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2) \quad (9)$$

and

$$J(\omega) = S^2\tau_m/(1 + \omega^2\tau_m{}^2) + (S_f{}^2 - S^2)\tau/(1 + \omega^2\tau^2) \quad (10)$$

where the order parameter $S^2$ is expressed as the product of two order parameters, $S_f^2$ and $S_s^2$, describing the degree of spatial restriction of the intramolecular motion of the internuclear $^{15}N-^1H$ vector internal motions on a fast and slow time scale, respectively. $\tau_m$ represents the overall tumbling correlation time of the entire molecule. The effective correlation time $\tau_e$ is expressed as $1/\tau = 1/\tau_e + 1/\tau_m$, and for the extended formalism of Clore et al. (*41*, *42*), when $S_f$ is $<1$, $\tau_e$ results from the slower internal motions $\tau_s$, where $1/\tau = 1/\tau_s + 1/\tau_m$. However, when $\tau_e$ is much smaller than $\tau_m$, the formalism is simplified to

$$J(\omega) = S^2\tau_m/(1 + \omega^2\tau_m{}^2) \quad (11)$$

Equation 11 can be used to estimate the overall correlation time $\tau_m$ from the ratio of $T_1/T_2$, which is 20.0 for Genesis, averaged from residues in well-structured sequences (*41*−*44*). Computer simulation is used to determine the most appropriate spectral density function and extract dynamic parameters of each residue. In the simulation, the target function $\chi^2$ was minimized and was defined as

$$\chi^2 = (T_{1c} - T_{1e})^2/\sigma_{T_1}{}^2 + (T_{2c} - T_{2e})^2/\sigma_{T_2}{}^2 + (\mathrm{NOE_c} - \mathrm{NOE_e})^2/\sigma_{\mathrm{NOE}}{}^2 \quad (12)$$

where the subscripts c and e represent calculated and experimentally determined relaxation parameters, respectively. $\sigma_{T_1}$, $\sigma_{T_2}$, and $\sigma_{NOE}$ are the standard deviations of the experimental data. Nonlinear least-squares optimization and Monte Carlo simulations were performed to minimize $\chi^2$ in the analyses as described previously (*35*, *45*).

From an average backbone structure determined from NMR constraints (*46*), the ratio of three principal components of Genesis (residues 1−93) is estimated to be 1.0:0.95:0.70. The motions of proteins with similar ratios were successfully described by the model free formalism with the assumption that the proteins tumbled isotropically in solution (*29*, *45*). Therefore, in our analysis, the overall motion of Genesis is assumed to be isotropic. Starting from an estimated $\tau_m$ (14.1 ns) which was determined from the $T_1/T_2$ ratio averaged (*41*−*43*) from structured regions in Genesis, the following models were employed for the analysis of the experimentally determined relaxation data: (i) $S^2$ using the simplified formalism (eq 11), (ii) $S^2$ and $\tau_e$ and (iii) $S^2$, $\tau_e$, and $R_{ex}$ using the standard formalism (eq 9), and (iv) $S_f^2$, $S_s^2$, and $\tau_e$ using the extended formalism (eq 10). In models i and ii, $\tau_m$ was further optimized for Genesis.

## RESULTS

*Analysis* of *Relaxation Parameters*. The intensities of individual amide resonances were measured from the 2D spectra taken at 290 K. The expressed DNA binding domain of Genesis contains 108 amino acids, which include an extra Met at the N terminus for the translation initiation, LQHHHHHH at the C terminus for Ni-NTA affinity purification (Quiagen), and 9 prolines in the sequence. Due to the highly overlapped nature of the amide signals in DNA free Genesis, only 67 of 107 singly protonated amides were well resolved for unambiguous peak height determinations under current experimental conditions (Figure 1). The signals from the C-terminal sequence and the $(His)_6$ tag of Genesis are highly overlapped and intense and have no assignment available; thus, no relaxation parameters can be obtained for residues after Leu93 (R94−H108). However, the 67 well-resolved $^1H-^{15}N$ peaks are distributed in all secondary structural elements of Genesis, and their dynamic properties can be viewed as being representative of the corresponding secondary structure (Figures 1 and 2). Genesis contains four helices and three sheets (Figure 2), and residues in these structured sequences show more uniform relaxation properties: $^{15}N$ $T_1$ in the range of 1.1−1.2 s and $^{15}N$ $T_2$ in the range of 50−60 ms. A $\tau_m$ of 14.5 ns was first estimated from the average $T_1/T_2$ ratio from the residues in structured regions. This unusually large $\tau_m$ is partly due to the relatively low temperature (290 K) at which the NMR data were acquired. The residues in wing 1 and wing 2 show slightly reduced $T_1$ values, increased $T_2$ values, and reduced NOE values. Thus, judged on the relaxation parameters alone, the two wings are expected to have large amplitude internal motions.

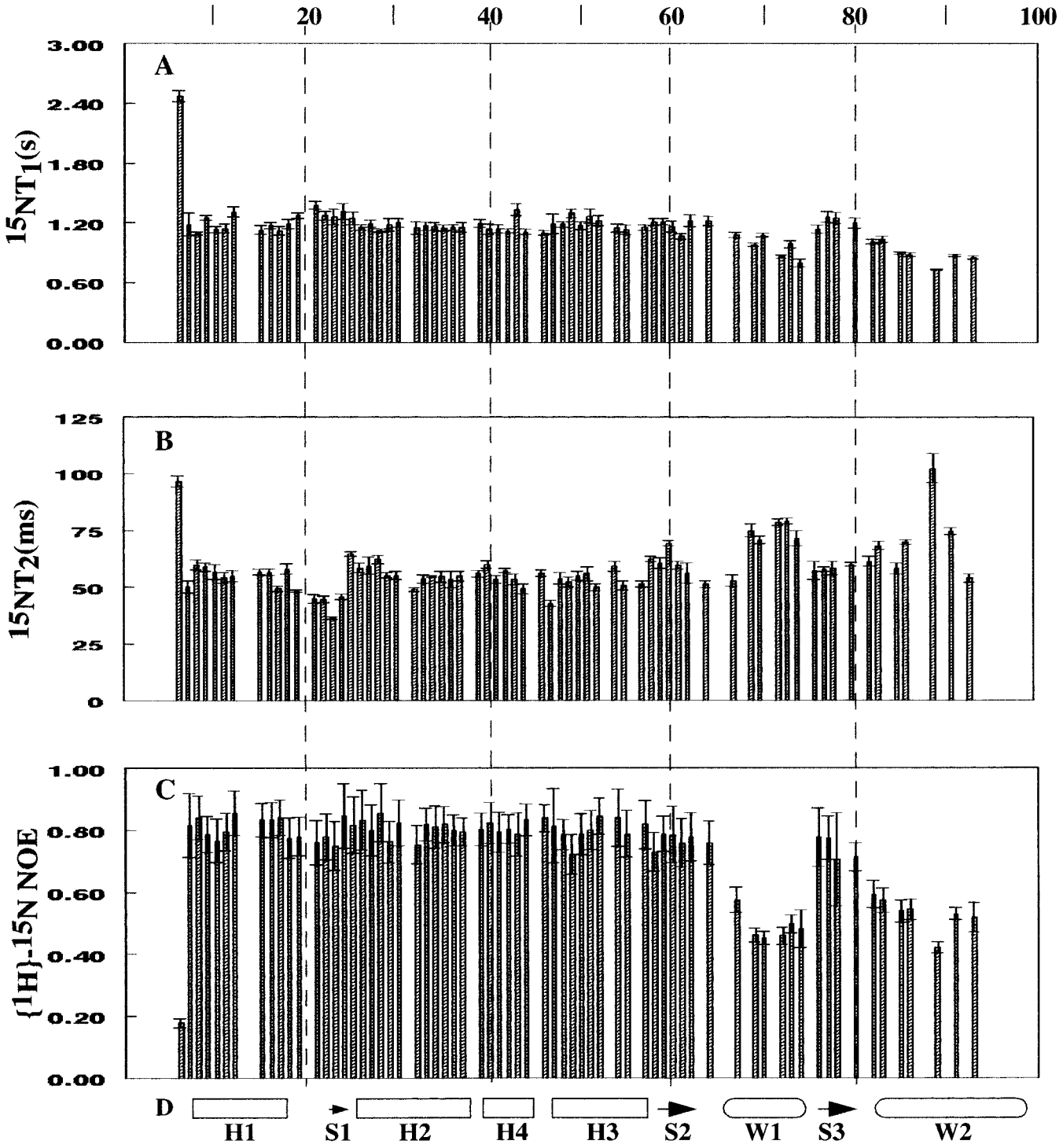


FIGURE 1: Relaxation parameters of Genesis acquired on a Bruker DRX 600 instrument at 290 K: (A) $^{15}$N $T_1$ with error, (B) $^{15}$N $T_2$ with error, (C) $^1$H$-^{15}$N steady-state NOE with error, and (D) the secondary structure of Genesis.

*Genesis Requires More Than One $\tau_m$ To Describe Overall Motions of Its Single Domain.* Model free analysis was performed on Genesis using the relaxation parameters that were obtained (Figure 1). Genesis is a single-domain polypeptide and binds to DNA as a monomer. The internal motions of most of the residues can be described by either the simplified formalism (eq 11) or the standard formalism (eq 9) with a single optimized $\tau_m$ of 13.9 ns (Figure 3). Although the relaxation parameters indicate that the two wings are likely to have large amplitudes of internal motions, the internal motions of most residues in wing 1 and wing 2 cannot be described by any form of the model free formalisms with 95% fidelity (Table 1), when a $\tau_m$ of 13.9 ns was used (*39*−*41*). One possibility is that the time scales and amplitudes of motions of the two wings are slow and large, and these motions are comparable with the overall motions of Genesis under our experimental conditions. In this case, these internal motions would directly modify the apparent $\tau_m$ for residues in the two wings. Therefore, separate $\tau_m$ values were further optimized for these two stretches of amino acid sequence by using the standard formalism (eq 9). Our data indicate that the optimized $\tau_m$ value for residues Gly69 and Gly74 is 11.5 ns and for residues Asn70, Gly73, Lys73, and residues starting from Glu82 is 10.9 ns. By using

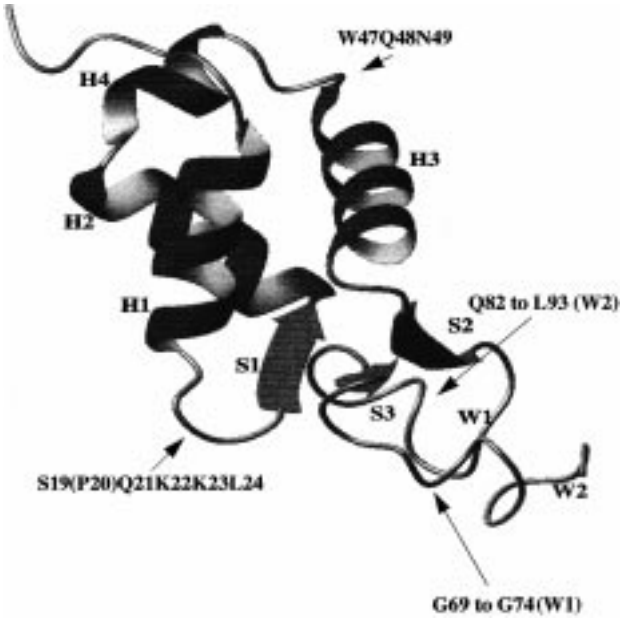


FIGURE 2: Ribbon diagram showing the tertiary structure of Genesis. Due to the lack of assignment for the C terminus of Genesis and NMR constraints, the calculated structure contains only residues 1−93. This structure was obtained from NMR-derived constraints with the program Diana (*47*) and was deposited in the Protein Data Bank (PDB code 2hfh). The secondary structures are as follows: H1, T8−Q18; S1, K23−T25; H2, L26−R36; H4, P38−F44; H3, W47−L57; S2, F61−I64; W1, P65−N75; S3, Y76−N80; and W2, P81−L93. The sequences which show sequence specific collective motions in our model free analysis are also indicated.

Table 1: Summary of $\chi^2$ Values from the Model Free Analyses of Wing 1 and Wing 2

| residue | $\chi^{2\,a}$ | $\chi^{2\,b}$ | $\chi^{2\,c}$ |
|---|---|---|---|
| G69 | 63.6 | 0.0 | 3.7 |
| N70 | 30.3 | 33.6 | 3.9 |
| G72 | 314.2 | 326.1 | 3.2 |
| K73 | 68.8 | 68.8 | 3.3 |
| G74 | 41.0 | 42.5 | 4.2 |
| Q82 | 10.4 | 6.8 | 4.6 |
| S83 | 67.8 | 22.2 | 5.8 |
| D85 | 113.6 | 113.6 | 4.4 |
| M86 | 165.8 | 165.8 | 6.5 |
| N89 | 352.8 | 352.8 | 6.0 |
| S91 | 217.0 | 0.0 | 3.1 |
| L93 | 68.4 | 68.4 | 6.4 |

[a] $\chi^2$ from an analysis using the standard model free formalism (eq 9). [b] $\chi^2$ from an analysis using the extended formalism (eq 10). In the first two analyses, the $\tau_m$ value was fixed at 13.9 ns. [c] $\chi^2$ values representing 95% fidelity, which can be reached in the analyses using the standard formalism (eq 9) with smaller $\tau_m$ values.

these different $\tau_m$ values, the internal motions in the residues between Gly69 and Gly74 and after Glu82 can be described by the standard model free formalism (eq 9) with 95% fidelity. This result indicates that there are relative motions on a nanosecond time scale between the two wings and the rest of the structured sequences in the DNA binding domain of Genesis. The relative motions between these secondary structures yield faster apparent overall motions for the residues in wing 2 and wing 1.

*The Sequence Connecting Helix 1 and Helix 2 Shows Motions on the Millisecond Time Scale*. Conformational exchanges are a good indicator of slow backbone motions on a millisecond time scale (*37*, *48*). The model free analysis of Genesis indicates that two stretches of sequence display large conformation exchange. One of the sequences is Ser19-(Pro20)-Glu21-Lys22-Lys23-Leu24. These residues link helix 1 and helix 2 and contain a turn and a $\beta$-strand (S1). In this sequence, residues show conformational exchanges ranging from 5.4 to 12.0 Hz except Pro20 for which no relaxation data are available (Figure 3). This result indicates that this stretch of amino acids in Genesis has collective motions on the millisecond time scale (*37*).

*Dynamics Analysis Supports the Existence of Helix 4*. Although Genesis is highly homologous to HNF-3$\gamma$, the two proteins demonstrate folding differences. An extra helix (H4) is observed in Genesis at a sequence (Figure 2) which is determined to be a loop in the crystal structure of HNF-3$\gamma$ (*23*, *24*). This fourth helix is inserted between helix 2 and helix 3 in Genesis (*24*, *46*). The insertion of this helix also yields a helix 3 in Genesis that is shorter than what was observed in HNF-3$\gamma$. Interestingly, the sequence forming helix 4 has been proposed to regulate DNA binding specificity in different winged helix family members, even though this sequence does not contact DNA directly (*27*). Our dynamic data strongly support the formation of helix 4 (*24*). In our model free analysis, residues 39−44 which form helix 4 in Genesis are relatively rigid, the average $S^2$ value is about 0.8, with no significant conformational exchange or internal motions described by $\tau_e$. Thus, this sequence shows a more structured nature in Genesis. Interestingly, residues Trp47-Glu48-Asn49 immediately following helix 4 show conformational exchange in the model free analysis. This result indicates that these three residues likely form the beginning of helix 3 or a loop connecting helix 3 and helix 4. In the structure of HNF-3$\gamma$, this conserved sequence Trp47-Glu48-Asn49 is near the center of helix 3.

*Several Potential DNA Contact Residues Have Conformation Exchange*. Residues Tyr6, Asn47, Leu56, and Trp76 are potential DNA contact residues, on the basis of amide chemical shift changes in a Genesis−DNA complex (C. Jin and X. Liao, unpublished). In our model free analysis, these residues also show conformational exchange (Figure 3). These potential motions on the millisecond time scale may play important roles in Genesis−DNA interactions.

## DISCUSSION

Protein−DNA interaction in transcription regulation is a highly dynamic process. A protein−DNA complex forms or dissociates in response to cell regulation signals. We have demonstrated that the DNA binding domain of the transcription factor Genesis is highly dynamic and that Genesis shows multiple autocorrelation times and sequence specific conformational exchange. These sequence specific collective motions are likely to be important in Genesis−DNA interactions.

Since a winged helix DNA binding domain is usually flanked by activation domains on both sides, the longest $\tau_m$ (13.9 ns) which represents the overall motion of Genesis under our experimental conditions is not likely meaningful in vivo. However, the relative motions between secondary structures are more likely to be preserved and to be important in Genesis function. Our data clearly show that the two DNA contact wings have collective motions on a nanosecond time scale relative to the helices and $\beta$-strands. These slow

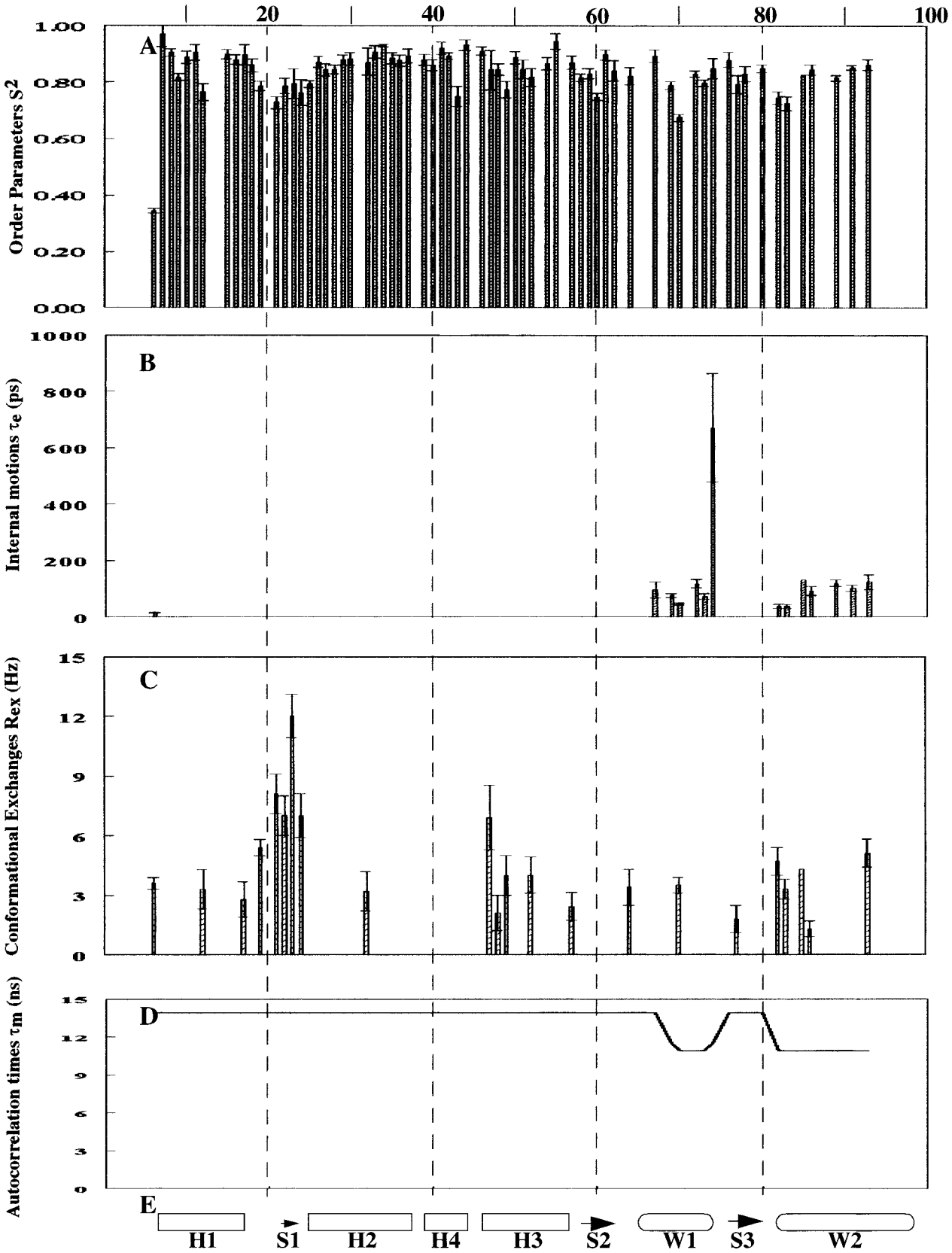


FIGURE 3: Model free analysis of Genesis: (A) $S^2$, (B) $\tau_e$, (C) $R_{ex}$, (D) multiple $\tau_m$ values needed for a description of the overall motions of Genesis, and (E) the secondary structure of Genesis.

motions associated with the two wings are likely to be independent of the overall motion of Genesis, and the smaller apparent $\tau_m$ values are the result of the addition of two motions in Genesis. In classical dynamics, the addition of two velocities follows the formulas $\mathbf{v}_0 = \mathbf{v}_1 + \mathbf{v}_2$, and $v_{0j} = v_{1j} + v_{2j}$ in the case where the projections of the vectors on axis $j$ are considered. Alternatively, $v_{0j} = v_{1j} + v_{2j}$ can be expressed as $1/\tau_{0j} = 1/\tau_{1j} + 1/\tau_{2j}$, in which $\tau_{ij}$ ($i = 0$, 1, or 2) is the time required to travel a unit distance on axis $j$ at speed $v_{ij}$ ($i = 0$, 1, or 2). Similarly, the apparent $\tau_m$ ($\tau_{mapp}$)

can also be assumed to contain two components, and if these two motions were isotropic and independent of each other, the $\tau_{mapp}$ would follow the formula

$$1/\tau_{mapp} = 1/\tau_m + 1/\tau_{mrel} \quad (13)$$

where $\tau_{mapp}$ is the apparent $\tau_m$ obtained for the mobile wings, $\tau_m$ is the overall correlation time of Genesis obtained from the structured sequence, and $\tau_{mrel}$ is the relative motion of the flexible sequences. In our study, the $\tau_m$ for Genesis is 13.9 ns, and $\tau_{mapp}$ for residues Asn70, Gly72, Lys73 and the residues after Glu82 is 10.9 ns. Thus, $\tau_{mrel}$ representing the relative motions of these sequences is 50.5 ns. Similarly, since Gly69 and Gly74 have a $\tau_{mapp}$ of 11.5 ns, the $\tau_{mrel}$ for these two residues is 66.6 ns. Thus, wing 1 and wing 2 have collective motions with a correlation time of around 50 ns. Apparently, these collective motions of wing 1 and wing 2 are not isotropic and cannot be fully described by the model free formalism. Nevertheless, $\tau_{mrel}$ can still give an estimate for the amplitudes of relative motions, and these relative motions are likely to be more important for protein−DNA interaction.

In our study, the residues after Leu93 in the Genesis DNA binding domain show strong and overlapped signals, and we expect these residues to be highly flexible under our experimental conditions. However, the motions of the N and C termini, which connect to activation domains, should be more restricted and should have smaller amplitudes in vivo.

Previous data showed that many DNA binding proteins and their cognate DNA sites adopt new conformations in protein−DNA complexes (*1*, *49*−*55*). On the basis of the structural and thermodynamic data, an "induced fit" model was proposed (*56*). In this model, the local folding transition of a DNA binding protein is coupled to site specific DNA binding. Then one question is whether some amino acid sequences in a DNA binding protein are more structurally flexible so that they can adopt alternative conformations in protein−DNA interactions. A clear answer to this question cannot be deduced easily from structural and thermodynamic data. However, this structural flexibility should be related to dynamic properties of a DNA binding protein.

Many structural motifs have been identified in protein−DNA interactions, some containing single-domain constructs and others containing multidomain constructs. The structures of many protein−DNA complexes have been determined by NMR and X-ray crystallographic techniques. These structures show that the formation of a stable complex usually involves amino acid residues from several structural elements of the protein in the complex. Thus, the conformational change of a DNA binding protein leads to rearrangement of the relative positions of the DNA contact residues on these structural elements. One such example is the $(Cys)_2(His)_2$ zinc finger DNA binding motif (*57*). Previous data showed that multifinger constructs were necessary for sequence specific DNA recognition, and that each of the fingers make DNA contacts (*57*, *58*). Although the structure of each zinc finger domain is rigid due to the zinc coordination, relative motions among the fingers have been observed in solution due to more flexible linker sequences (*59*, *60*). The relative positions of the DNA contact residues on different fingers are rearranged constantly due to the interfinger motions. Therefore, the conformations of multi-zinc finger proteins are highly flexible. However, many DNA binding motifs are single-domain DNA binding motifs. How a single-domain DNA binding protein adjusts its conformation in the DNA recognition process is less obvious. An extreme case is presented in the NMR studies of HMG domains. Two of the DNA sequence specific HMG domains, SRY and LEF-1, do not adopt defined structures before they contact DNA (*61*, *62*). In DNA recognition, SRY and LEF-1 have to recognize the linear DNA sites first and then adopt new conformations, which causes DNA site bending in DNA complexes. However, the DNA sequence nonspecific HMG domain HMG1, which has defined structure before it binds to DNA (*63*), binds preferentially to distorted DNA structures (*64*−*66*) and DNA with an enforced bend (*67*). Thus, it seems that the sequence flexibility associated with these less structured DNA free HMG proteins is required due to large conformational changes when they bind to DNA.

Our data indicate that collective motions are associated with the two DNA contacting wings in Genesis and that these collective motions are on the nanosecond time scales as indicated by corresponding $\tau_{mrel}$ values. These motions show Genesis constantly adjusting the relative positions of its DNA contact residues on the different secondary structure elements. On this basis, Genesis can be treated as a multidomain DNA binding protein; the four helices and three sheets compose one domain, and the two wings form the other two DNA contact domains. Furthermore, the two sequences between helix 1 and helix 2 and between helix 4 and helix 3 show considerable conformational exchange, which indicate motions on the millisecond time scale. Therefore, our data show that Genesis is highly dynamic and contains sequence specific collective motions. Previous data also show that Genesis recognizes both linear and prebent DNA sites in which the DNA molecules adopt many different conformations (*25*). Therefore, it is likely that the sequence specific collective motions in Genesis provide the motional freedom and flexibility for this unique DNA binding property. Indeed, our preliminary data obtained on a Genesis−DNA complex indicate that the structural and dynamic properties of these mobile sequences in DNA free Genesis are severely perturbed.

The winged helix DNA binding motif is a highly conserved DNA binding motif. Some conserved residues are difficult to understand if only the structures are considered. One such example is the sequence of wing 1, which contains many conserved prolines and glycines. Due to these prolines and glycines, wing 1 has no commonly observed structure (*23*, *24*). In this study, wing 1 of Genesis shows collective motions before contacting DNA. Thus, it seems that these residues in wing 1 are conserved to support a unique dynamic property in the winged helix proteins.

In this study, the $\tau_m$ value of Genesis is 13.5 ns at 290 K. This value is larger than expected for a protein of 108 amino acid residues. However, the $\tau_m$ estimated from the average $T_1/T_2$ ratio acquired at 295 K reduces to 8.5 ns which is more reasonable for Genesis at that temperature. Furthermore, no large chemical shift perturbations have been observed in spectra taken at the two temperatures. Thus, the large $\tau_m$ value observed at 290 K for Genesis is likely a result of low temperature and high concentration of salts used for the NMR

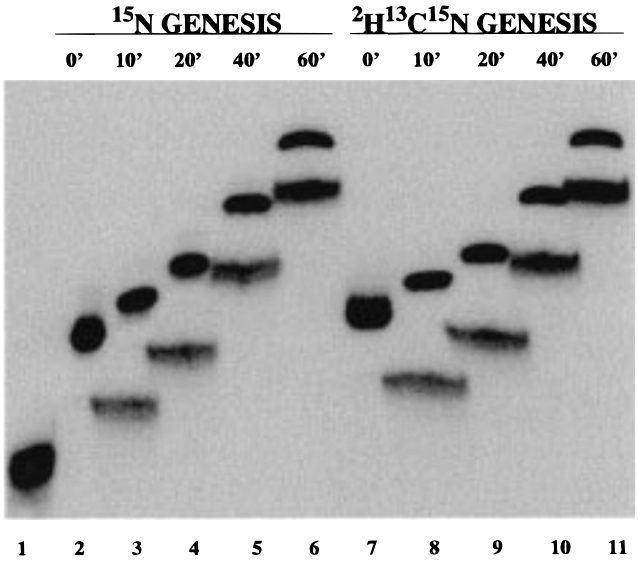


FIGURE 4: Gel shift assay for studying the possible isotope effects in the off rate of Genesis−DNA complexes. Lane 1 is $^{23}$P-labeled DNA probe #HFH-2#12. Lanes 2−6 show a time-dependent dissociation of the [$^{15}$N]Genesis−DNA complex in the presence of a 25-fold molar excess of the unlabeled #HFH-2#12 DNA site. In each lane, the faster migrating band is the $^{32}$P-labeled free probe after the unlabeled probe is added to the reaction. Lanes 7−11 show a time-dependent dissociation of the $^{15}$N-, $^{2}$H-, and $^{13}$C-enriched Genesis−DNA complex in the presence of a 25-fold molar excess of the unlabeled #HFH-2#12 DNA site.

study. This possibility is supported by a previous dynamic study of RNase H at different temperatures (*69*).

The Genesis protein used in this study was perdeuterated on all backbone and side chain carbon atoms. One of the advantages of using a perdeuterated protein is that the removal of $^{3}J_{HNC}$ coupling and $H^N-H^C$ cross relaxation increases the signal intensities dramatically, while reducing the line width in the $H^N$ dimension. Therefore, even though a relatively low temperature (290 K) was used, the data acquired from perdeuterated Genesis has a better signal-to-noise ratio, reduced overlap, and better statistical reliability. At this temperature, due to a long $\tau_m$, protonated Genesis shows low signal-to-noise ratio and large statistical uncertainties. Thus, deuterated proteins are better for dynamic studies on biological molecules with large autocorrelation times. Due to large uncertainties in the dynamic parameters acquired for protonated Genesis, the dynamic parameters for each individual residue cannot be extracted with high fidelity under the current experimental conditions. Perdeuteration of a protein increases its molecular weight by 10−15%, but the $\tau_m$ values for the perdeuterated protein (13.9 ns) and protonated Genesis (13.5 ns) are close. This is not surprising, since a protein is highly hydrated and is surrounded by a large volume of structured or semistructured water (*69*). These protein-associated $H_2O$ molecules are expected to rotate with the protein in solution; thus, a slightly increased molecular weight without an increase of the molecular volume of a protein will not slow the rotational motion dramatically. Furthermore, on the basis of this analysis, we expect that the large amplitude local motions associated with a surface residue should not be influenced by deuteration if this residue is hydrated. However, one other question is whether deuteration can modify the DNA binding property of Genesis due to the isotope effect of the heavier side chains and backbone. Protein−DNA interaction is a highly dynamic process, and the complex:free DNA binding protein ratio is determined by the on and off rate between a DNA binding protein and its cognate DNA site in the complex. The dissociation constant $K_d$ is the ratio of the off rate to the on rate. Under our current experimental conditions, we cannot measure the on rate of Genesis binding to DNA. However, an experiment comparing the off rate of triply labeled Genesis and [$^{15}$N]Genesis was performed (Figure 4). Our data indicated that the isotopic effect due to $^{2}$H and $^{13}$C enrichment in the side chains and backbone of Genesis has minimal influence on the off rate and binding affinity in the Genesis−DNA complex. Thus, our data indicate that perdeuteration provides a powerful tool for studies of protein dynamics.

## ACKNOWLEDGMENT

We thank Dr. R. H. Costa for the generous gift of the Genesis gene. We thank B. Curtiss for providing a macro for peak height determination. We thank Dr. A. Palmer for the program of the model free analysis. We thank Dr. P. W. Gettins for critical comments during preparation of the manuscript. We also thank J. Rebello for reading the manuscript.

## SUPPORTING INFORMATION AVAILABLE

One table giving the values and uncertainties, for each residue, of the experimentally determined $T_1$, $T_2$, and NOE data, the optimized order parameters, the internal correlation times and $^{15}$N exchange broadening terms, and overall correlation times (4 pages). Ordering information is given on any current masthead page.

## REFERENCES

1. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
2. Harrison, S. C. (1991) *Nature 353*, 715−719.
3. Raumann, B. E., Brown, B. M., and Sauer, R. T. (1994) *Curr. Opin. Struct. Biol. 4*, 36−43.
4. Kaufmann, E., and Knöchel, W. (1996) *Mech. Dev. 57*, 3−20.
5. Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H., and Darnell, J. E., Jr. (1990) *Genes Dev. 4*, 1427−1436.
6. Weigel, D., and Jäckle, H. (1990) *Cell 63*, 455−456.
7. Grossniklaus, U., Pearson, R. K., and Gehring, W. J. (1992) *Genes Dev. 6*, 1030−1051.
8. Zhu, G., Muller, E. G. D., Amacher, S. L., Northrop, J. L., and Davis, T. N. (1993*) Mol. Cell. Biol. 13*, 1779−1787.
9. Li, C., Lai, C., Sigman, D. S., and Graynor, R. B. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 7739−7743.
10. Tao, W., and Lai, E. (1992) *Neuron 8*, 957−966.
11. Clevidence, D. E., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E., and Costa, R. H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3948−3952.
12. Avraham, K. B., Fletcher, C., Overdier, D. G., Clevidence, D. E., Lai, E., Costa, R. H., Jenkins, N. A., and Copeland, N. G. (1995) *Genomics 25*, 388−393.
13. Bassel-duby, R., Hernandez, M. D., Yang, Q., Rochelle, J. M., Sledin, M. F., and Williams, R. S. (1994) *Mol. Cell. Biol. 14*, 4596−4605.
14. Dirkson, M. L., and Jamrich, M. (1992) *Genes Dev. 6*, 599−608.
15. Miller, L. M., Gallegos, M. E., Morisseau, B. A., and Kim, S. K. (1993) *Genes Dev. 7*, 933−947.
16. Pierrou, S., Hellqvist, M., Samuelsson, L., EnerbÇck, S., and Carlsson, P. (1994) *EMBO J. 13*, 5002−5012.

17. Costa, R. H., Grayson, D. R., and Darnell, J. E., Jr. (1989) *Mol. Cell. Biol. 9*, 1415−1425.
18. Costa, R. H., and Grayson, D. R. (1991) *Nucleic Acids Res. 19*, 4139−4145.
19. Liu, E., DiPersio, C. M., and Zaret, K. S. (1991) *Mol. Cell. Biol. 11*, 773−784.
20. Ang, S. L., and Rossant, J. (1994) *Cell 78*, 561−574.
21. Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, R. M., and Darnell, J. E., Jr. (1994) *Cell 78*, 575−588.
22. Ye, H., Kelly, T. F., Samadani, U., Lim, L., Rubio, S., Overdier, D. G., Roebuck, K. A., and Costa, R. H. (1997) *Mol. Cell. Biol. 17*, 1626−1641.
23. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature 364*, 412−420.
24. Marsden, I., Chen, Y., Jin, C., and Liao, X. (1997) *Biochemistry 36*, 13248−13255.
25. Bravieri, R., Shiyanova, T., Chen, T. H., Overdier, D., and Liao, X. (1997) *Nucleic Acids Res. 25*, 2888−2896.
26. Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, D. A., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., Klemsz, M., and Hromas, R. (1996) *J. Biol. Chem. 271*, 23126−23133.
27. Overdier, D. G., Porcella, A., and Costa, R. H. (1994) *Mol. Cell. Biol. 14*, 2755−2766.
28. Garner, M., and Rezin, A. (1981) *Nucleic Acids Res. 9*, 6505−6525.
29. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
30. Sklenar, V., Torchia, D. A., and Bax, A. (1987) *J. Magn. Reson. 73*, 375−379.
31. Boyd, J., Hummel, U., and Campbell, I. D. (1990) *FEBS Lett. 175*, 477−482.
32. Palmer, A. G., Skelton, N. J., Chazin, W. J., Wright, P. E., and Rance M. (1992) *Mol. Phys. 75*, 699−711.
33. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
34. Markley, J. L., Horsley, W. J., and Klein, M. P. (1971) *J. Chem. Phys. 55*, 3604−3605.
35. Palmer, A. G., Wright, P. E., and Rance, M. (1991) *J. Am. Chem. Soc. 113*, 203−216.
36. Abragam, A. (1961) *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, U.K.
37. Bloom, M., Reevers, L. W., and Wells, E. J. (1965) *J. Chem. Phys. 42*, 1615−1642.
38. Hiyama, Y., Niu, C., Silverton, J. V., Bavoso, A., and Torchia, D. A. (1988) *J. Am. Chem. Soc. 110*, 2378−2383.
39. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4558.
40. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
41. Clore, G. M., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.
42. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989−4991.
43. Key, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.
44. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) *Biochemistry 31*, 5269−5278.
45. Stone, M. J., Fairbrother, W. J., Palmer, A. G., III, Reizer, J. J., Saier, M. H., Jr., and Wright, P. E. (1992) *Biochemistry 31*, 4394−4406.
46. Marsden, I., Jin, C., and Liao, X. (1998) *J. Mol. Biol.* (in press).
47. Günter, P., Braun, W., and Wüthrich K. (1991) *J. Mol. Biol. 217*, 517−530.
48. Palmer, A. G., III (1993) *Curr. Opin. Biotechnol. 4*, 385−391.
49. Harrison, S. C., and Aggarwal, A. K. (1990) *Annu. Rev. Biochem. 59*, 933−969.
50. Werner, M. H., Gronenborn, A. M., and Clore, G. M. (1996) *Science 271*, 778−784.
51. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science 253*, 1001−1007.
52. Rosenberg, J. M. (1991) *Curr. Opin. Struct. Biol. 1*, 104−132.
53. Winkler, F. K. (1993) *EMBO J. 12*, 1781−1795.
54. Weber, I. T., and Steiz, T. A. (1987) *J. Mol. Biol. 198*, 311−326.
55. Brennan, R. G. (1991) *Curr. Opin. Struct. Biol. 1*, 80−104.
56. Spolar, R., and Record, T., Jr. (1994) *Science 263*, 777−784.
57. El-Baradi, T., and Pieler, T. (1991) *Mech. Dev. 35*, 155−169.
58. Liao, X., Clemens, K. R., Tennant, L., Wright, P. E., and Gottesfeld, J. M. (1992) *J. Mol. Biol. 223*, 857−871.
59. Bruschweiler, R., Liao, X., and Wright, P. E. (1995) *Science 268*, 886−889.
60. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) *Nat. Struct. Biol. 4*, 605−608.
61. Werner, M. H., Huth, J. R., Groneenborn, A. M., and Clore, G. M. (1995) *Cell 2*, 705−714.
62. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) *Nature 376*, 791−795.
63. Hardman, C. H., Broadhurst, W., Raine, A. R. C., Grasser, K. D., Thomas, H. O., and Laue, E. D. (1995) *Biochemistry 34*, 16596−16607.
64. Bianchi, M. E., Beltrame, M., and Paonessa, G. (1989) *Science 243*, 1056−1059.
65. Pil, P. M., and Lippard, S. J. (1992) *Science 256*, 234−237.
66. Locker, D., Decoville, M., Maurizot, J. C., Bianchi, M. E., and Lend, M. (1995) *J. Mol. Biol. 246*, 243−247.
67. Wolfe, S. A., Ferentz, A. E., Grantcharova, V., Churchill, M. E. A., and Verdine, G. L. (1995) *Chem. Biol. 2*, 213−221.
68. Mandel, A. M., Akke, M., and Palmer, A. G., III (1996) *Biochemistry 35*, 16009−16023.
69. Cantor, C., and Schimmel, P. R. (1980) *Biophysical Chemistry*, p 550, Academic Press, New York.

BI980031V